

Searching Lattice Data Structures of Varying Degrees of Sortedness *

Mohammad Obiedat

Department of Science, Technology, and Mathematics,
Gallaudet University, 800 Florida Avenue NE, Washington,
DC 20002-3695, USA[†]

May 16, 2016

Abstract

Lattice data structures are space efficient and cache-suitable data structures. The basic searching, insertion, and deletion operations are of time complexity $O(\sqrt{N})$. We give a jump searching algorithm of time complexity $O(J(L) \log(N))$, where $J(L)$ is the jump factor of the lattice. $J(L)$ approaches 4 when the degree of sortedness of the lattice approaches \sqrt{N} . A sorting procedure of time complexity $O(\sqrt{N})$ that can be used, during the system idle time, to increase the degree of sortedness of the lattice is given.

Keywords. Lattice data structure, cache-suitable, jump searching, jump factor, degree of sortedness, incremental sorting.

1 Introduction

Searching is one of the most time-consuming operation of many software applications. Basically, a searching procedure takes a target key and it has to

*Published in the International Journal of Data Structures, 1(1) 64-76 (2015)

[†]Email address: mohammad.obiedat@gallaudet.edu

decide whether that key is present or absent. This seemingly simple operation becomes very challenging when the set of keys is dynamically changing, by insertion and deletion, or when using minimum space and energy are very crucial. Therefore, any data structure that is designed to organize pieces of information in a dynamic environment should be evaluated by the space required to hold the keys, by the time required to perform the searching, insertion, and deletion operations, and by the suitability of the structure for caching or more generally for memory hierarchies.

In Bjarne Stroustrup’s recent paper [12], he emphasizes the importance of compactness and predictable memory access patterns for efficiency. Stroustrup points out that power consumption is roughly proportional to the number of memory accesses. Consequently, the less data we store and move the better for efficiency, especially in infrastructure software and applications for hand-held devices like smartphones. Stroustrup makes a strong case in favor of using arrays instead of linked lists or more generally pointer-based structures. There are two important differences between arrays and linked lists that affect the efficiency of their operations. First, arrays are more space efficient than linked lists. Second, arrays are more suitable for caching than linked lists, because an array’s elements are stored in contiguous memory locations, while a linked list’s nodes are scattered in memory.

Most of the current index structures are either pointer-based or use extra storage space in order to obtain the optimal $O(\log(N))$ time complexity of their operations. Obviously, such data structures are not suitable for memory-constrained systems. Sequential search of unsorted arrays is space efficient and very suitable for caching, but unfortunately it is of time complexity $O(N)$, which is too slow for most applications.

The lattice data structure (LDS) was novelly introduced by Berkovich in 1992, see [2]. The space required to hold a lattice with N keys is equal to the space required to hold an array with N keys. Additionally, it is possible to devise more efficient algorithms using the LDS than using the array. This makes the LDS a good contribution to the current trends of developing data structures and algorithms for memory-constrained applications. Also, the LDS operations have good temporal and spatial locality which makes it a rich environment for the ongoing research on cache-conscious data structures and cache-oblivious algorithms. Consequently, in addition to the theoretical importance of the LDS, we argue that this structure is the ideal choice for memory-constrained systems, for real-time systems, for systems where power consumption is crucial, and for systems where incremental adjustment is

feasible.

The basic searching, insertion, and deletion operations of the LDS are of time complexity $O(\sqrt{N})$. In this paper, we give a general jump searching algorithm of time complexity $O(J(L) \log(N))$, where $J(L)$ is the jump factor of the lattice. The jump factor of the lattice approaches 4 when the degree of sortedness of the lattice approaches \sqrt{N} . In order to keep the jump factor of the lattice small, we provide a sorting procedure of time complexity $O(\sqrt{N})$ that can be used, during the system idle time (e.g., by giving it the least priority when scheduling the lattice operations), to increase the degree of sortedness of the lattice.

Experimental evaluation of randomly built LDSs shows that the performance of the basic searching procedure of the LDS is similar to skip list search when the structures have up to 1000 keys, and the performance of the jump searching procedure is similar to skip list search when the degree of sortedness of the lattice is close to $0.9h$, where h is the height of the lattice. When the ratio of the number of times the sorting procedure is called to the number of times the insertion and deletion operation are called is close to 5, the performance of the jump searching procedure is similar to its performance on a lattice with degree of sortedness close to $0.9h$.

To obtain a good performance of the searching operation, a hybrid searching algorithm can be implemented, together with the sorting procedure, by searching the lattice using the basic searching procedure when the degree of the sortedness of the lattice is less than or equal to $0.9h$, and the jump searching procedure when the degree of the sortedness of the lattice is greater than $0.9h$.

The layout of this paper is as follows. In Section 2, we formalize the definition of the LDS, present the basic searching, insertion, and deletion algorithms, and then analyze the space and time complexities of implementing these algorithms. In Section 3, we devise a general jump searching algorithm that can be used instead of the basic searching algorithm to obtain a better performance and time complexity when the jump factor of the lattice is small. Then, we provide a detailed analysis of the jump factor and devise a sorting algorithm that can be used to reduce the jump factor of the lattice during the system idle time. In Section 4, we compare the performance of the jump searching algorithm with skip lists search. Then, we examine the effect of using the sorting algorithm on the jump factor of the lattice during the system idle time. Conclusion and recommendations for further research are given in Section 5.

2 Lattice Data Structures

In this section, we formally define the lattice data structure and discuss its basic algorithms and properties.

In order to define the lattice data structure, we first recall the definition of Ferrers diagrams, see [13]. Let $\lambda = (n_1 \geq \dots \geq n_m > 0)$ be an integer partition of a positive integer N . A Ferrers diagram of shape λ consists of m rows, where the top row contains n_1 equisized square cells (hereafter referred to as cells), the second row from the top contains n_2 cells, etc. Each row is left-justified. For a given non-negative integer number h , we are interested in up-side-down Ferrers diagrams of shape $(h+3, h+2, \dots, 2, 1)$, see Figure 1 for an up-side-down Ferrers diagram of shape $(7, 6, 5, 4, 3, 2, 1)$. The following numbering schema and terminologies will be used for such diagrams. Rows and columns are numbered from bottom to top and from left to right, starting with one, respectively. Cells in each row and each column are numbered from left to right and from bottom to top, starting with one, respectively. By diagonal k , we mean the cells that lie on the line segment that can be formed by connecting the upper-left corner of the first cell of row k and the lower-right corner of the first cell of column k where $1 \leq k \leq h+3$. The head of diagonal k is the first cell of row k and the tail is the first cell of column k . Cells in each diagonal are numbered from head to tail, starting with one. Using up-side-down Ferrers diagrams, instead of the usual Ferrers diagrams, is only a matter of convenience and preference.

Definition 2.1 *A lattice data structure (LDS) of height h is an arrangement of a finite set of distinct positive integers, referred to as proper keys, in an up-side-down Ferrers diagram of shape $(h+3, h+2, \dots, 2, 1)$ such that:*

1. *Each cell in row one contains 0.*
2. *Each cell in column one contains 0.*
3. *Each cell in diagonal $h+3$ contains ∞ , except the head and the trail which contain 0.*
4. *All other cells contain proper keys, except the last k non-zero cells of diagonal $h+2$ which contain ∞ for some $0 \leq k \leq h-1$.*
5. *The proper keys in each row, column, and diagonal are in increasing order from left to right, from bottom to top, and from head to tail, respectively.*

Remark 2.2 *In the original definition of the LDS given by Berkovich [2], diagonals were not required to be sorted. As we will see in the next section, by requiring diagonals to be sorted one can apply a variety of searching techniques to improve the time complexity of the basic searching algorithm.*

Unless otherwise indicated, we will not distinguish between a cell's location and its content, so we will say that “cell two in row four is less than cell three in row four” instead of “the content of cell two in row four is less than the content of cell three in row four”. A LDS of height 4 and proper keys $\{3, 5, 6, 9, 12, 20, 30, 31\}$ is shown in Figure 1.

0							
0	∞						
0	12	∞					
0	9	30	∞				
0	5	20	∞	∞			
0	3	6	31	∞	∞		
0	0	0	0	0	0	0	0

Figure 1: A LDS of height 4

Remark 2.3 *The purpose of the improper keys in the boundary cells, row one, column one, and diagonal $h+3$, is to serve as a wall; the LDS procedures can not pass these cells and move down, to the left, or to the right. Row one cells will serve as failure cells; if a searching procedure reaches these cells then it means that the key is absent. Choosing the keys to be positive integers and the relation \leq to be the usual less than or equal relationship is only a matter of convenience; in practice the keys could belong to any set with a partial order relationship.*

In Figure 2, we illustrate the eight possible movements from a given cell C to an adjacent cell. These movements will be used in devising the LDS algorithms.

UL	U	UR
L	C	R
DL	D	DR

U = Up, D = Down, L = Left, R = Right

UL = Up Left, UR = Up Right

DL = Down Left, DR = Down Right

Figure 2: Movements in a LDS from a given cell C

Remark 2.4 *If the cells adjacent to C preserve the lattice structure, i.e., satisfy condition (5) of Definition 2.1, then C also preserves the lattice structure provided that $C > D$, $C > UL$, $C < U$, and $C < DR$ or $DR = 0$.*

Next, we describe the basic searching, insertion, and deletion algorithms of the LDS. We assume that we are given a LDS of height $h \geq 1$ and a key K in the search space of the lattice.

Searching. The basic searching algorithm of the LDS, SearchLDS, works as follows. First, we set C to be the second cell of diagonal $h + 2$. Then we perform the following movements until K is declared present or absent: If $K > C$, then we move diagonally to DR by setting $C = DR$. If $K < C$, then we move down to D by setting $C = D$. We declare K is present if $C = K$, and declare K is absent if $C = 0$. This basic searching algorithm, SearchLDS, is given as Algorithm 1.

Insertion. To insert a key K into a LDS, we first search the lattice for K . If K is present then it cannot be inserted. If K is absent, then we find a temporary cell for K as follows: If diagonal $h + 2$ has cells containing ∞ , then we insert K at the first cell of diagonal $h + 2$ that contains ∞ . On the other hand, if diagonal $h + 2$ does not have any cell that contains ∞ , then we expand the lattice by increasing its height by one and then insert K in the second cell of diagonal $h + 3$. After K is inserted in the temporary cell, say cell C , we maintain the lattice structure by applying the inward algorithm, Inward, given as Algorithm 2.

Algorithm 1 Basic Searching Algorithm of the LDS

```
SearchLDS (LDS,  $h$ ,  $K$ )
// Given a LDS of height  $h$ , and a key  $K$ .
// Determine whether  $K$  is present or absent.
 $C$  = second cell of diagonal  $h + 2$ ;
while ( $C \neq K$  and  $C \neq 0$ ) do
    if ( $K > C$ ) then
         $C = DR$ ; // a  $DR$  movement
    else
         $C = D$ ; // a  $D$  movement
    end if
end while
if  $C = K$  then
     $K$  is present at  $C$ ;
else
     $K$  is absent;
end if
```

Deletion. To delete a key K , we first find its location, say K is located at cell C . If C is the last proper key of diagonal $h + 2$, then we place ∞ in cell C . If C is not the last proper key of diagonal $h + 2$, then we place the last proper key in diagonal $h + 2$, say F , in cell C and place ∞ in cell F . Finally, we maintain the lattice structure by applying the Inward algorithm if ($C < D$ or $C < UL$), and the Outward algorithm, given as Algorithm 3, if ($C > U$ or $C > DR$ and $DR \neq 0$). In all cases, we reduce the height of lattice by one if diagonal $h + 2$ does not have any proper key.

In the following theorem, we give the number of proper and improper keys of a LDS of height h .

Theorem 2.5 *Let L be a lattice of height h , where $h \geq 1$, and suppose that diagonal $h + 2$ contains k proper keys where $1 \leq k \leq h$. Then*

1. *The number of proper keys in L is $h(h - 1)/2 + k$.*
2. *The number of improper keys in L is $4h - k + 6$.*

Proof: Straightforward.

Algorithm 2 Inward Algorithm of the LDS

Inward (LDS, C)
while ($C < D$ **or** $C < UL$) **do**
 if ($UL > D$) **then**
 swap UL and C ; // a UL swap
 else
 swap D and C ; // a D swap
 end if
end while

Algorithm 3 Outward Algorithm of the LDS

Outward (LDS, C)
while ($C > U$ **or** ($C > DR$ **and** $DR \neq 0$)) **do**
 if ($DR < U$ **and** $DR \neq 0$) **then**
 swap DR and C ; // a DR swap
 else
 swap U and C ; // a U swap
 end if
end while

Corollary 2.6 *The smallest lattice that can hold N proper keys is of height h where $h = \lfloor (1 + \sqrt{8N - 7})/2 \rfloor = \lceil (-1 + \sqrt{8N + 1})/2 \rceil$.*

In the following lemma, we give upper bounds on the number of comparisons to search for a key, and the average number of comparisons to search for a present key when a LDS is searched using SearchLDS.

Lemma 2.7 *Let L be a lattice of height h and suppose that diagonal $h + 2$ contains k proper keys where $1 \leq k \leq h$. If we search L using SearchLDS, then*

1. *The number of comparisons to search for any absent key is $h + 1$.*
2. *The number of comparisons to search for any present key is less than or equal to h .*
3. *The average number of comparisons to search for a present key is*

$$\frac{2h^3 - 2h + 3k^2 + 3k}{3h^2 - 3h + 6k} \approx \frac{2h}{3}.$$

Proof: Straightforward.

Theorem 2.8 *The time complexity of SearchLDS, Inward, and Outward algorithms of a LDS with N proper keys is $O(\sqrt{N})$.*

Proof: From Lemma 2.7 and Corollary 2.6, the time complexity of SearchLDS is $O(\sqrt{N})$. To find the time complexity of the Inward algorithm, suppose C is located on diagonal s and column t . Then with each D swap, s is reduced by 1 and t is not changed, while with each UL swap, t is reduced by 1 and s is not changed. So with each swap, $s + t$ is reduced by 1. Hence the Inward algorithm makes up to $2h$ swaps. Consequently, it is an $O(\sqrt{N})$ algorithm. Similarly, the time complexity of the Outward algorithm is $O(\sqrt{N})$.

Corollary 2.9 *If SearchLDS is used to locate a key or determine its absence, then the time complexity of the insertion and deletion procedures, described in this section, of a LDS with N proper keys is $O(\sqrt{N})$.*

One of the main advantages of LDS over other searching structures such as balanced trees (e.g., AVL trees [11] and self-adjusting trees [10]), hash tables [4], skip lists [6, 7, 8], Jumplists [3], and J-lists [1] is the space required to hold and maintain the lattice. A lattice can be easily represented, either explicitly or implicitly, by a one-dimensional array where each cell is represented by one array element. In explicit representation of LDSs, a lattice is mapped into a one-dimensional array by mapping diagonal one to the first array element, diagonal two to the second and third array elements, etc. Implicit representation of LDSs is similar to explicit representation, except that only cells with proper keys are mapped into array elements. While implicit mapping is a little more space efficient than explicit mapping, explicit mapping is the natural choice, because of its simplicity and its superiority in term of time.

In Table 1, we compare the extra space required to hold a proper key by various searching data structures.

3 Jump Searching of Lattice Data Structures

Lattice data structures outperform other data structures such as balanced trees, self-adjusting trees, and skip lists in terms of space, time bounds of some operations, and suitability for memory hierarchies. However, the basic

Table 1: Required extra space per proper key by various searching structures

Structure	Extra space per proper key
LDSs-Explicit mapping	approaches 0 as $h \rightarrow \infty$
LDSs-Implicit mapping	0
J-lists	1 integer
Skip lists	an average of 4/3 pointers
Jumplist	2 pointers and 1 integer
AVL trees	2 pointers and 2 bits
Red-black trees	2 pointers and 1 bit
Splay trees	2 pointers
Hash tables	variable

searching algorithm of the LDS, of time complexity $O(\sqrt{N})$, cannot compete with the searching algorithms of these data structures in terms of time, especially when the size of the search space is large. In this section, using the fact that diagonals and columns of LDSs are sorted, we devise a jump searching algorithm that can be used instead of SearchLDS to speed up the search. The basic idea of the jump searching algorithm is to make wide jumps on diagonals and columns instead of the sequential movements to adjacent cells in the basic searching algorithm. We also devise a sorting algorithm that can be used, during the system idle time, to incrementally increase the degree of sortedness of the lattice, and hence to improve the performance of the jump searching algorithm.

Jump searching algorithms over portions of a sorted array are studied in detail in [9]. In Algorithm 4, JumpSearchLDS, we given a general jump searching algorithm of the LDS that can be used with a variety of jump searching procedures of sorted arrays (e.g., binary and interpolation) to obtain a better performance and time complexity than SearchLDS.

Unless otherwise indicated, from now on, we will assume that L is a LDS of height $h \geq 1$, and K is a key in the search space of L . For brevity, let d stand for a DR movement. We define the search path of K , with respect to L , to be the sequence of movements (d or D) until SearchLDS terminates.

Definition 3.1 *The jump factor of K , with respect to L , is the number of blocks of identical movements in the search path of K .*

We denote the jump factor of K with respect to L by $J(K; L)$, or $J(K)$, when L is understood. For example, if the search path of K is $ddddDDDDdd$

Algorithm 4 Jump Searching Algorithm of the LDS

```
JumpSearchLDS (LDS,  $h$ ,  $K$ )
// Given a LDS of height  $h$ , and a key  $K$ .
// Determine whether  $K$  is present or absent.
 $C$  = second cell of diagonal  $h + 2$ ;
while ( $C \neq K$  and  $C \neq 0$ ) do
    if ( $K < C$ ) then
        // a downward jump
        Let  $C$  be the first cell in the current column such that  $K \geq C$ ;
    else
        // a diagonal jump
        Let  $C$  be the first cell in the current diagonal such that  $K \leq C$ , set
         $C = 0$  if there is no such cell;
    end if
end while
if  $C = K$  then
     $K$  is present at  $C$ ;
else
     $K$  is absent;
end if
```

then $J(K) = 3$. We define the jump factor of L by

$$J(L) = \max\{J(K) : K \text{ is a key in the search space of } L\}.$$

Let P be the procedure that performs the jump steps in JumpSearchLDS. So, P will take a sorted array, a column (respectively, a diagonal) with m elements and for a given key K will return the first element F such that $K \geq F$ (respectively, the first element F such that $K \leq F$ or set $C = 0$ if there is no such F). In the following theorem, we give the time complexity of JumpSearchLDS where the time complexity of P is $O(t_P(m))$.

Theorem 3.2 *Let L be a lattice of height h . If the time complexity of P is $O(t_P(m))$, then the time complexity of JumpSearchLDS is $O(J(L)t_P(h))$.*

Proof: Given a lattice of height h . One can form a triangle by connecting the following three points: the lower-left corner of the first cell of row one, the lower-right corner of the last cell of row one, and the upper-left corner of the

last cell of column one. In Algorithm 4, when a downward jump is performed the keys in the area outside triangle T_1 in Figure 3-A are eliminated from the search, and when a diagonal jump is performed the keys in area outside triangle T_2 in Figure 3-B are eliminated from the search. Since the time complexity of each jump is $O(t_P(h))$ and each search requires up to $J(L)$ jumps, then the time complexity of JumpSearchLDS is $O(J(L)t_P(h))$.

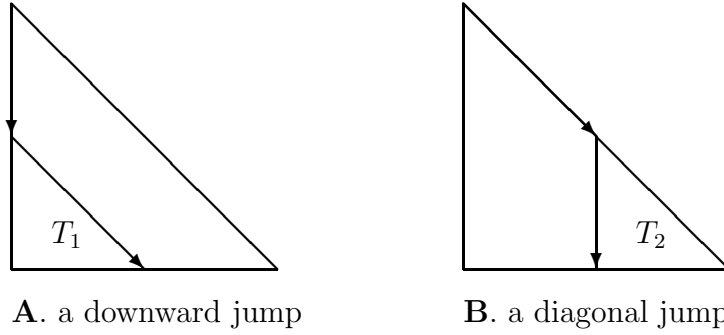


Figure 3: Jumps of the LDS

Corollary 3.3 *Let L be a lattice with N proper keys. If the time complexity of P is $O(\log(m))$, then the time complexity of JumpSearchLDS is $O(J(L)\log(N))$.*

The time complexity of JumpSearchLDS depends on the jump factor of the lattice and on the procedure used to perform the jumps. From Theorems 2.8 and 3.2, if a binary-search-like procedure is used to perform the jump steps, then the time complexity of JumpSearchLDS is less than or equal the time complexity of SearchLDS provided that the jump factor of the lattice is less than or equal to $\sqrt{N}/\log(N)$. In the following lemma, we give upper bounds on the jump factors of keys.

Lemma 3.4 *Let L be a lattice of height h , and K be a key in the search space of L . Suppose the last movement in the search path of K ends on diagonal s and column t . Let $u = t - 2$ and $v = h - s + 2$, then*

1. *The number of d movements in the search path of K is u and the number of D movements is v .*

2. $J(K) \leq \begin{cases} u + v & \text{if } u = v \\ 2u + 1 & \text{if } u < v \\ 2v + 1 & \text{if } u > v \end{cases}.$
3. $J(K) \leq \begin{cases} h - 1 & \text{if } K \text{ is present} \\ h & \text{if } K \text{ is absent} \end{cases}.$

Proof: Straightforward.

Theorem 3.5 *Let L be a lattice of height h . Then*

1. *The average of the jump factors of the present keys is less than or equal to $h/3 + 5/6 \approx h/3$.*
2. *The average of the jump factors of the absent keys is less than or equal to $h/2 + 1/2 \approx h/2$, provided that the probability the search for a randomly chosen absent key will terminate in a cell on row 1 and column t for each $2 \leq t \leq h + 2$ is $1/(h + 1)$.*

Proof: Follows from Lemma 3.4 and by using basic algebraic summations.

Definition 3.6 *Suppose $\alpha \in \{3, \dots, h\}$. A LDS is said to be sorted of degree α (for brevity, α -sorted) if the first proper key of diagonal s is greater than the last proper key of diagonal $s - 1$ for each $4 \leq s \leq \alpha + 2$.*

For example, the lattice in Figure 1 is 3-sorted but it is not 4-sorted.

Remark 3.7 *If L is α -sorted, then the proper keys in the first $\alpha + 2$ diagonals are completely sorted. Also, if $\beta = \alpha/h$ then the percentage of the proper keys in the first $\alpha + 2$ diagonals is $\beta^2 + (\beta - \beta^2)/(h + 1) \approx \beta^2$. For example, if $\alpha = h/2$ then approximately 25% of the proper keys in the lattice are completely sorted.*

Lemma 3.8 *Suppose L is α -sorted and K is a key in the search space of L .*

1. *If K is present and the last movement in the search path of K ends on diagonal s , where $s < \alpha + 2$, then $J(K) \leq 2$.*
2. *If K is absent and the last movement in the search path of K ends on diagonal s , where $s < \alpha + 1$, then $J(K) \leq 4$.*

Proof: (1) Suppose K is present and $s < \alpha + 2$. Since the first proper key in diagonal $\alpha + 2$ is greater than the last proper key in diagonal $\alpha + 1$, then K is less than the first proper key in diagonal $\alpha + 2$. So, by a diagram chasing, the search path of K is either $D \cdots D$ or $D \cdots Dd \cdots d$. Hence, $J(K) \leq 2$.
(2) Suppose K is absent and $s < \alpha + 1$. By a diagram chasing, the search path of K is either $D \cdots D$, $D \cdots Dd \cdots d$, $D \cdots Dd \cdots dD$, or $D \cdots Dd \cdots dDd \cdots d$. Hence, $J(K) \leq 4$.

Theorem 3.9 *If L is α -sorted, then $J(L) \leq \max\{4, 2h - 2\alpha + 2\}$.*

Proof: Follows directly from Lemma 3.8.

Corollary 3.10 *Let L be an α -sorted lattice of height h . If the time complexity of the procedure that performs the jump steps in *JumpSearchLDS* is $O(t_P(m))$, then the time complexity of *JumpSearchLDS* is $O(\max\{4, 2h - 2\alpha + 2\}t_P(h))$.*

Corollary 3.11 *If L is h -sorted, then*

1. $J(K) \leq 2$ for any present key K , and
2. $J(K) \leq 4$ for any absent key K .

Corollary 3.12 *Let L be an h -sorted lattice with N proper keys. If the time complexity of the procedure that performs the jump steps in *JumpSearchLDS* is $O(\log(m))$, then the time complexity of *JumpSearchLDS* is $O(\log(N))$.*

In the following theorem, we give upper bounds on the average of the jump factors of the present and absent keys for α -sorted lattices.

Theorem 3.13 *Suppose L is α -sorted and $\beta = \alpha/h$. If $0.5 \leq \beta \leq 1$, then*

1. *The average of the jump factors for the present keys is less than or equal to*

$$\frac{(2\beta^3 - 4\beta^2 + 2\beta)h^2 - (3\beta^2 - 6\beta + 1)h + \beta - 3}{h + 1}$$

$$\approx (2\beta^3 - 4\beta^2 + 2\beta)h - (3\beta^2 - 6\beta + 1).$$

2. *The average of the jump factors for the absent keys is less than or equal to*

$$\frac{(\beta^2 - 2\beta + 1)h^2 + 4h - 4}{h + 1} \\ \approx (\beta^2 - 2\beta + 1)h + 4,$$

provided that the probability the search for a randomly chosen absent key will terminate in a cell on row 1 and column t for each $2 \leq t \leq h+2$ is $1/(h+1)$.

Proof: To prove this theorem, we only need to use Lemmas 3.4, 3.8, and basic algebraic summations, then substitute $\alpha = \beta h$.

Remark 3.14 *The functions $f(\beta) = 2\beta^3 - 4\beta^2 + 2\beta$ and $g(\beta) = \beta^2 - 2\beta + 1$ are decreasing on $[0.5, 1]$ with $f(0.5) = g(0.5) = 0.25$ and $f(1) = g(1) = 0$. So, from Theorem 3.13, when β approaches 1, the average of the jump factors for the present keys approaches 2 and the average of jump factors for the absent keys approaches 4. Also, from Theorem 3.9, $J(L) \leq \max\{4, (2 - 2\beta)h + 2\}$.*

From Theorem 3.13, the average of the jump factors of a lattice data structure varies inversely with the degree of sortedness of the lattice. So, in order for JumpSearchLDS to compete with searching structures of time complexity $O(\log(N))$, we need to maintain LDS with high degrees of sortedness. In Algorithm 5, SortLDS, we provide a procedure for incrementally increasing the degrees of sortedness of LDS. SortLDS can be used during the system idle time to increase the degree of sortedness of the lattice.

In the following theorem we give the time complexity of SortLDS.

Theorem 3.15 *The time complexity of SortLDS algorithm of a LDS with N proper keys is $O(\sqrt{N})$.*

Proof: Straightforward.

Algorithm 5 Sorting Algorithm of the LDS

```
SortLDS( $T, h$ )  
  // Given a LDS  $T$  of height  $h$ .  
  // Swap the keys in the first two cells that prevent  $T$  from being  $h$ -sorted.  
   $i = 4$ ;  
  while ( $i \leq h + 1$  and the last proper key on diagonal  $i$  is less than the  
    first proper key on diagonal  $i + 1$ ) do  
     $i = i + 1$ ;  
  end while  
  if  $i = h + 2$  then  
    break; // The lattice is  $h$ -sorted  
  else  
    let  $C$  be the second cell of diagonal  $i + 1$   
    let  $F$  be the first cell of diagonal  $i$  with key larger than the key in  $C$ ;  
    swap the keys in  $F$  and  $C$ ;  
    Outward( $T, C$ );  
  end if
```

4 Experimental Evaluation of LDSs and Skip Lists

In this section, we examine the performance of the LDS searching algorithms for lattices with varying jump factors. We compare the time performance values of SearchLDS, JumpSearchLDS, and skip lists search. At the end of the section, we examine the effect of using SortLDS on the jump factor of the lattice during the system idle time.

A skip list is a probabilistic searching data structure, invented by William Pugh in the late 1980's [6, 8]. Skip lists have most of the desirable properties of balanced trees such as AVL trees and self-adjusting trees with $O(\log(N))$ average time for most operations, yet they are more simple and space efficient. The rationale for comparing LDSs with skip lists is twofold. First, skip lists are among the newest searching data structures, they are compared against AVL trees and self-adjusting trees in [8], so by comparing LDSs with skip lists we gain insight into the relationship between these important structures and LDSs. Second, the most important advantages of LDSs are their simplicity and space efficiency, but these are claimed to be the most important advantages of skip lists too.

The LDS algorithms are implemented in C programming language using Microsoft Visual Studio 2010 Compiler. Skip lists source C code is obtained from William Pugh by anonymous ftp to ftp.cs.umd.edu. The proper keys are randomly chosen from the set $[1, \text{RAND_MAX}] = [1, 2147483647]$ and stored in a file. Then the file is used to build and search lattices of varying heights and jump factors. Same keys are used to build and search skip lists. A binary-search-like procedure is used to perform the jump steps in JumpSearchLDS. The time performance values are the CPU time, measured in microseconds, on Intel(R) Core(TM) i5-2400 CPU @ 3.10GHz with 4.0GB RAM computer. Carrying out the experiments using the GNU C compiler on a Unix environment shows similar comparable results.

From Lemma 3.4, the number of each of the d and D movements in the search path for any key depends only on the location where the search path terminates. So, the performance of SearchLDS is independent of the jump factor of the lattice. In Table 2, we compare the performance of SearchLDS, skip lists search, and JumpSearchLDS for lattices with varying jump factors. The time is the average searching time of the present keys, h is height of the lattice, and N is the number of keys. As we see from the table, the performance of JumpSearchLDS is similar to SearchLDS when $\beta = 0.8$, JumpSearchLDS starts to perform better than SearchLDS when $\beta \geq 0.9$, that is when more than 81% of the keys are completely sorted. On the other hand, the performance of JumpSearchLDS is similar to skip lists search when $\beta = 0.9$, JumpSearchLDS starts to perform better than skip lists search when $\beta \geq 0.95$, that is when more than 90% of the keys are completely sorted. When $\beta = 1$ the lattice is completely sorted and searching using JumpSearchLDS is similar to searching a sorted array using binary search.

Table 2: Timings of SearchLDS, skip lists search, and JumpSearchLDS

h	N	SearchLDS	Skip List	JumpSearchLDS			
				$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 1.0$
10	55	0.048	0.059	0.064	0.061	0.061	0.060
50	1275	0.190	0.144	0.215	0.155	0.134	0.128
100	5050	0.347	0.207	0.370	0.223	0.176	0.157
500	125250	1.840	0.481	1.686	0.678	0.357	0.228
1000	500500	3.990	0.780	3.600	1.284	0.557	0.267

If the lattice is not h -sorted, then \sqrt{N} calls to SortLDS increase the degree

of sortedness of the lattice by at least 1, and hence reduce the average of the jump factors of the present keys of the lattice. So, using SortLDS during the system idle time will incrementally adjust the lattice, by increasing its degree of sortedness, and hence improves the performance of JumpSearchLDS.

In order to simulate the performance of SortLDS during the system idle time, we will assume the following scenario: the system starts with an h -sorted lattice with N proper keys, when the system is not idle the operations search, insert, and delete are performed, when the system is idle SortLDS is repeatedly called, the system goes through idle time when $0.1N$ new keys are inserted, and the ratio of the insert operations to the delete operations is equal to 2. The purpose of this hypothetical scenario is to give us a rough idea about the impact of using SortLDS, during the system idle time, on the average of the jump factors of the present keys of the lattice. Let γ to be the ratio of the number of times SortLDS is called to the number of times the insert and delete operations are called. Since the time complexity of the insert, delete, and SortLDS procedures is \sqrt{N} , then, in Table 3, instead of calculating the average of the jump factors of the present keys of the lattice with specific ratios of the idle time to the time of insert and delete operations, we will calculate the average of the jump factors of the present keys for the lattice with varying values of γ .

Table 3: Average of the jump factors of the present keys for incrementally adjusted lattices using SortLDS

h	Average jump factor with varying γ				
	$\gamma = 0$	$\gamma = 3$	$\gamma = 4$	$\gamma = 5$	$\gamma = 6$
10	2.2	1.82	1.77	1.77	1.77
50	5.16	3.86	2.95	2.11	1.96
100	8.03	5.81	4.14	2.75	1.97
500	29.11	20.47	14.27	8.02	2.43
1000	52.47	37.46	25.66	13.87	3.11

In Table 4, we calculate the average of the jump factors of lattices of varying heights and degrees of sortedness.

From Tables 3, 4, the average of the jump factors of the present keys for incrementally adjusted lattices using SortLDS with $\gamma = 0$, respectively $\gamma = 5$, is close to the average of the jump factors of the present keys for sorted lattices with $\beta = 0.8$, respectively $\beta = 0.90$. So, using the above scenario, when the

Table 4: Average of the jump factors of the present keys for lattices with varying degrees of sortedness

h	Average jump factor with varying β			
	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.95$	$\beta = 1.0$
10	1.93	1.75	1.64	1.64
50	4.45	2.70	2.05	1.92
100	7.26	3.53	2.38	1.96
500	29.53	10.22	4.26	1.99
1000	57.37	18.57	6.54	2.00

system starts $\beta = 1.0$, and searching using JumpSearchLDS is similar to searching a sorted array using binary search. As the operations insert and delete are performed, β decreases from 1.0 to 0.8 and JumpSearchLDS slows down and becomes similar to SearchLDS. Using SortLDS with $\gamma = 5$ increases β from 0.8 to 0.90, and searching using JumpSearchLDS becomes similar to skip lists search.

5 Conclusions

In this paper, we have formalized the definition of the lattice data structure and presented its basic operations. We have shown that the LDS is as space efficient as the array and it is very suitable for caching. The worst case time complexity of the LDS basic operations is better than the corresponding worst case time complexity of most of other data structures operations. In order to make this structure more attractive in term of the average case time complexity of the searching procedure, we have devised a general jump searching algorithm of time complexity $O(J(L) \log(N))$. We have also provided a sorting algorithm of time complexity $O(\sqrt{N})$ that can be used to reduce $J(L)$ during the system idle time. One possible direction for further research is to explore using randomized searching procedures where the search starts at different locations of the lattice instead of the second cell of diagonal $h + 2$. Another possible direction is to perform a comprehensive experimental comparison of the LDS with B^+ -Trees and T -Trees to find out the best suitable structure for indexing in main memory. In [5], we showed that the LDS is a very robust structure for the order-statistic operations.

Acknowledgements. I would like to thank Simon Berkovich for his helpful comments and suggestions. I am also thankful to Regina Nuzzo who carefully read the manuscript of this paper.

References

- [1] J. Bentley, F. Leighton, M. Lepley, D. Stanat, and J. Steele. A Randomized Data Structure for Ordered Sets. *Advances in Computing Research*, 5:413-428, 1989.
- [2] S. Berkovich. Organization of Associative Memory Operations With Lattice Structures. *Proceedings of the 35th IEEE Midwest Symposium On Circuits and Systems*, 2:887-890, 1992.
- [3] H. Bronnimann, F. Cazals, and M. Durand. Randomized Jumplists: A Jump-and-Walk Dictionary Data Structure. *Proceedings of the 20th Annual Symposium on Theoretical Aspects of Computer Science, LNCS*, Springer, 2607:283-294, 2003.
- [4] D. Knuth. *The Art of Computer Programming Vol. 3: Sorting and Searching*. Adison-Wesley Publishing Company, 1973.
- [5] M. Obiedat. Time-efficient Algorithms for the Order-statistic Operations in Lattice Data Structures. To be published.
- [6] W. Pugh. A Skip List Cookbook. *Computer Science Technical Report Series, CS-TR-2286.1*, 1-12, 1990.
- [7] W. Pugh. Concurrent Maintenance of Skip Lists. *Computer Science Technical Report Series, CS-TR-2222.1*, 1-13, 1990.
- [8] W. Pugh. Skip Lists: A Probabilistic Alternative to Balanced Trees, *LNCS 382 Springer*, 437-449 1989.
- [9] B. Shneiderman. Jump Searching: A Fast Sequential Search Technique. *Comm. of the ACM*, 21(10):831-834, 1978.
- [10] D. Sleator and R. Tarjan. Self-adjusting Binary Search Trees. *Comm. of the ACM*, 32(3):652-666, 1985.

- [11] S. Štrbac-Savić and M. Tomašević. Comparative Performance Evaluation of the AVL and Red-Black Trees, Proceedings of the Fifth Balkan Conference in Informatics, 14-19, 2012.
- [12] B. Stroustrup. Software Development for Infrastructure. Computer, 45(1):47-58, 2012.
- [13] A. Yong. What is a Young Tableau?. Notice of the AMS, 54(2):240-241, 2007.